

Procedure to construct a multi-scale coarse-grained model of DNA-coated colloids from experimental data

Bianca M. Mladek^{*a,b}, Julia Fornleitner^c, Francisco J. Martinez-Veracoechea^a, Alexandre Dawid^d, and Daan Frenkel^a

Received Xth XXXXXXXXXX 2010, Accepted Xth XXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXX 200X

DOI: 10.1039/b000000x

We present a quantitative, multi-scale coarse-grained model of DNA coated colloids. The parameters of this model are transferable and are solely based on experimental data. As a test case, we focus on nano-sized colloids carrying single-stranded DNA strands of length comparable to the colloids' size. We show that in this regime, the common theoretical approach of assuming pairwise additivity of the colloidal pair interactions leads to quantitatively and sometimes even qualitatively wrong predictions of the phase behaviour of DNA-grafted colloids. Comparing to experimental data, we find that our coarse-grained model correctly predicts the equilibrium structure and melting temperature of the formed solids. Due to limited experimental information on the persistence length of single-stranded DNA, some quantitative discrepancies are found in the prediction of spatial quantities. With the availability of better experimental data, the present approach provides a path for the rational design of DNA-functionalised building blocks that can self-assemble in complex, three-dimensional structures.

1 Introduction

In the pursuit of designing materials that self-assemble into specific target structures suitable building blocks have to be found with interactions that drive the formation of these structures. The availability of such tailor-made nano-structured materials could open the way to many interesting applications¹. In order to program self-assembly into nano-sized building blocks, it is crucial that the interactions between these building blocks can be tuned. One class of potential 'programmable' building blocks are colloidal particles functionalised with polymers. Such particles can be designed in many different shapes, ranging in sizes from nm to μm^2 . Moreover, the precise choice of their polymeric coating, i.e. type, length, flexibility, grafting density and architecture of the polymers, allows for additional freedom in tuning the interactions between the particles. Among such systems, DNA-coated colloids (DNACCs) have received special attention^{3–5}, mainly because the technology exists to produce specific DNA strands quickly and cheaply. These colloidal particles carry

short single-stranded (ss) DNA sequences ("sticky ends") connected to inert, grafted polymers ("spacers"). Three-dimensional aggregates of such colloids can then be formed due to the highly specific and temperature-reversible hybridisation of complementary sticky ends; these are either carried by different species of colloids or are part of so-called linker sequences that bridge between different colloids.

The aggregation behaviour of DNACCs can indeed be influenced via the properties of the colloids, their polymeric coating, as well as the solution in which the particles are immersed. Experiments of nano-^{6–11} and micron-sized^{12–14} DNACCs have shown that self-assembly of *simple* spatially ordered structures, such as bcc or fcc crystals, is possible. However, applications such as photonic band-gap materials would require non-close-packed crystals of low coordination, such as the diamond structure¹⁵, and despite recent progress in the field^{16–18}, the *design* of arbitrarily complex ordered structures is presently still challenging. One crucial factor is that DNACCs tend to assemble more readily into amorphous aggregates than into spatially ordered structures¹⁹. The reason is that the attractions between the DNACCs are strongly dependent on the external conditions such as ionic strength or temperature^{3,20,21}. Exquisite control over these parameters is thus needed to help DNACCs to anneal into ordered structures. Unless we improve our ability to design DNACCs that assemble readily into the desired target structure, the practical use of these building blocks remains limited.

It is for this reason that the use of coarse-grained models

^a Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom.

^b Department of Structural and Computational Biology, Max F. Perutz Laboratories GmbH, University of Vienna, Dr. Bohr-Gasse 9, 1030 Vienna, Austria. Tel: 0043 1 4277 52272 ; E-mail: bianca.mladek@univie.ac.at

^c Institute for Complex Systems, Forschungszentrum Jülich, 52428 Jülich, Germany

^d Université Joseph Fourier Grenoble 1/CNRS, Laboratoire Interdisciplinaire de Physique UMR 5588, Grenoble, 38041, France

is explored both in theoretical approaches^{21–25} and in computer simulations^{26–37}. These models allow for a fast and efficient exploration of new design principles of DNACCs. This opens the way to develop strategies for crystals to form in broader temperature windows³⁸ and offer greater freedom in the design of DNACCs and the structures they form^{14,24,25,28,30,34,39–41}.

Existing models typically range from highly simplified ones (e.g. lattice models^{27,40} or pair interaction approaches^{22,33}) to sophisticated models featuring explicit modelling of the DNA hybridisation^{26,29–32,34–37}. In addition, many models exploit the elastic properties of the DNA strands: very long strands can be described by scaling laws^{27,29}, while short strands of double-stranded (ds) DNA can be represented as rigid rods^{21,32,38,42}. Many of the existing models are qualitative and focus on the generic features of DNACC self-assembly – typically, these models do not aim to describe any specific DNACC system and hence do not exploit the full available experimental information about the building blocks.

However, for the computer-aided design of DNACCs, quantitative, but computationally tractable models of DNACCs are needed. By comparison to the qualitative models mentioned above, models of quantitative predictive power are rare^{42–45}, and they are most successful at describing micron-sized colloids covered with short dsDNA strands, where a description of the systems via pair interactions determined from simulations has proven successful. However, the DNACCs that have, thus far, shown most promise for crystallisation are the ones for which the radius of gyration of the ssDNA strands, R_g , is of comparable size to the radius of the colloids, R_C . In this regime, the modelling strategies that are successful for larger colloids cannot be applied: the strands are usually too flexible to be approximated as rigid rods but too short for polymer scaling laws to apply.

In a recent Letter³⁷ we showed that multi-stage coarse graining can be used to describe the phase behaviour of nano-sized DNACCs functionalised with ssDNA. The present manuscript describes in detail the methodology that we have developed to arrive at such a multi-stage coarse-grained model. The text is organised as follows: In the first three sections, we present three steps of coarse-graining in which we identify the key degrees of freedom that determine the phase behaviour of DNACCs: we develop our most detailed model of DNACCs based on experimental data in Sec. 2. Based on simulations of this model, we derive the “core-blob model”, the second level of coarse-graining, in Sec. 3. This model, in turn, allows us to perform the final step of coarse-graining and calculate effective interactions (Sec. 4). The expected reliability of the effective interactions to predict the phase behaviour of DNACCs is assessed in Sec. 5. Finally, we calculate the phase diagram of the chosen DNACCs within both the core-blob model and the effective interaction approach in Sec. 6.

In the Appendices, we detail technical aspects of the present work.

2 Stage 1: Model with explicit DNA chains

2.1 General outline

To develop the most detailed level of description of DNACCs, a suitable model for the ssDNA strands tethered to the colloid’s surface has to be chosen. ssDNA is not a simple polymer: it is prone to form hairpin structures and knots. An accurate description of such substructures could be achieved by fully atomistic simulations, which are computationally feasible at most for small systems of DNACCs covered with few, short DNA strands^{46,47}. Alternatively, rather detailed, coarse grained models of DNA such as developed in Ref. 48 could be employed. But while this model makes the detailed study of hybridisation between several strands of DNA feasible⁴⁹, it is still computationally too time-consuming to be employed for a system of hundreds or thousands of colloids, each covered with dozens of ssDNAs. Fortunately, the formation of ssDNA loops and knots is expected to play a minor role for DNACCs where commonly DNAs are chosen that are not self-complementary; we thus neglect this effect and model the ssDNA strands as freely jointed, charged chains. This model captures the most important contribution to the behaviour of the ssDNA strands which stems from the electrostatic repulsion of the DNA’s sugar-phosphate backbone⁵⁰. In view of the high Young’s modulus of ssDNA⁵⁰, the segments of every freely jointed, charged chain are chosen to have a fixed Kuhn length $l_{\text{Kuhn}} = 2p_{\text{ss}}$, where p_{ss} is the persistence length of ssDNA. The number of Kuhn segments n_K per chain is determined as

$$n_K = \left\lfloor \frac{l_{\text{contour}}}{l_{\text{Kuhn}}} \right\rfloor. \quad (1)$$

Here, $l_{\text{contour}} = (N_b - 1)b_0$ is the contour length of the ssDNA strand, with N_b the number of bases per strand, and b_0 the interbase distance in ssDNA. The symbol $\lfloor \dots \rfloor$ denotes the floor function. We stress that since $l_{\text{Kuhn}} > b_0$, each Kuhn segment represents several nucleotides. Consequently, our model cannot capture the precise base sequence of the ssDNA strands; sequence dependent effects, such as base stacking, are captured only in an averaged way in the choice of b_0 (see below).

The conformation of a freely jointed, charged chain is defined by the positions $\{\mathbf{r}_0, \dots, \mathbf{r}_{n_K}\}$ of the $n_K + 1$ vertices of the chain. We approximate the continuous charge of the backbone by effective charges sitting at each of these vertices; two vertices i and j at distance $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ interact with each other via a Debye-Hückel interaction, Φ_{ij}^{ele} , given by

$$\Phi_{ij}^{\text{ele}} = \frac{q^2}{4\pi\epsilon_0 D} \frac{e^{-\kappa r_{ij}}}{r_{ij}}, \quad (2)$$

where D is the dielectric constant of the solvent and ϵ_0 is the vacuum permittivity (in SI units). The charge per vertex, q , is approximated as

$$q = \frac{l_{\text{contour}}}{(n_K + 1)} v, \quad (3)$$

with the effective line charge density of ssDNA, v . Finally, the inverse Debye screening length, κ , in Eqn. 2 is given by

$$\kappa = \sqrt{\frac{2\beta I N_A e^2 10^3}{D\epsilon_0}}, \quad (4)$$

where N_A is Avogadro's number and e is the elementary charge. $\beta = 1/k_B T$, where k_B denotes Boltzmann's constant and T stands for the temperature (all in SI units). The ionic strength I of the solution in which the DNACCs are immersed is given in mol/l and the factor 10^3 stems from converting mol/l to SI units.

The n_K segments of each ssDNA strand can be divided into two classes: the number of sticky end segments $n_{\text{se}} = \lfloor (N_{b,\text{se}} - 1)b_0/l_{\text{Kuhn}} \rfloor$, calculated from the number of sticky end bases $N_{b,\text{se}}$; and $n_{\text{sp}} = n_K - n_{\text{se}}$, the number of spacer segments. A total of N_{str} ssDNA strands are then grafted to the colloid: we attach the DNA strands at their first vertex \mathbf{r}_0 , neglecting in our model the hexane-thiol group by which the ssDNA strands are experimentally tethered and which is estimated to have a end-to-end length of ~ 0.8 nm⁴⁷. The colloid, in turn, is modelled as a hard sphere of radius R_C that cannot be penetrated by the vertices of the DNA strands (see Fig. 1). In experiments^{7,51}, colloids are typically maximally loaded with DNAs and therefore we assume that the anchoring points are uniformly distributed on the surface of the colloid and—in accordance with experimental evidence⁵²—that they cannot diffuse.

2.2 Chosen values

Knowing the experimental conditions under which the reference experiments were performed, we can determine the values of all variables introduced in the last section. As a proof-of-concept of our method, we choose to study system V from Ref. 7, where a symmetric, binary mixture of gold nano-colloids (labelled A and B) of radius $R_C \sim 6$ nm were studied. All colloids were coated with ~ 60 ssDNA strands of $N_b = 65$ bases, out of which $N_{b,\text{se}} = 15$ bases constitute the sticky end. A and B colloids only differed in their sticky end sequences, which were complementary to allow for direct hybridisation between the unlike colloids. These DNACCs were assembled in a solution of 0.01 mol/l phosphate buffer, 0.2 mol/l NaCl at pH = 7.1⁷. Using the Henderson-Hasselbach equation⁵³, the concentration c_i of all ion species $i = 1, \dots, n$ in the solution can be calculated; then, the ionic strength, I , is given as, $I = \frac{1}{2} \sum_{i=1}^n c_i z_i^2$, where z_i denotes the charge number of

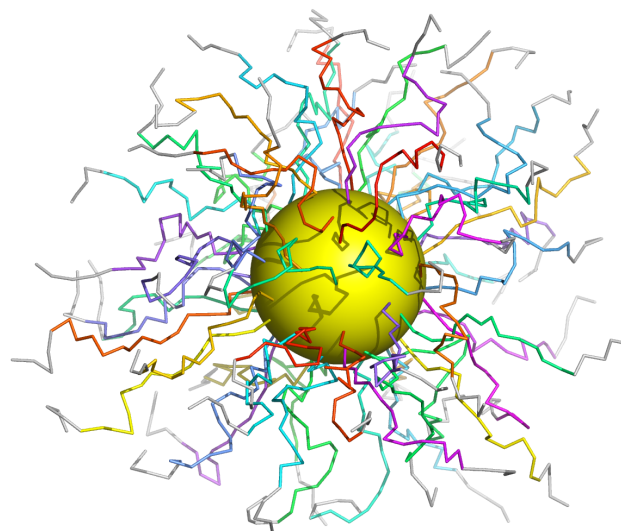


Fig. 1 (Colour online) Simulation snapshot of our most detailed model of DNACCs: a hard sphere colloid (yellow) of $R_C = 6$ nm is dressed with 60 strands of freely jointed, charged chains of 18 Kuhn segments (corresponding to 65 bases; various colours) at the solvent conditions given in Ref. 7 and at $T = 25^\circ\text{C}$. The 4 last segments model the sticky ends and are plotted in grey.

ion species i . For the present system⁷, we find that $I = 0.21$ mol/l.

Next, we need to set the persistence length of ssDNA. While the persistence length of dsDNA is well known, its value for ssDNA is less well established: a broad range of experimentally measured values has been published, varying between 0.75 nm and almost 10 nm⁵⁴. Here, we use a value of 0.75 nm, in accordance with the studies on which our ssDNA model is based⁵⁰. According to Ref. 55, p_{ss} depends on the ionic strength and the dependence is approximately described by $p_{\text{ss}}[\text{\AA}] \sim 4I[\text{mol/l}]^{-1/2}$. Therefore, at the ionic strength of our reference experiment (i.e. $I = 0.21$ mol/l), this results in $p_{\text{ss}} \sim 0.87$ nm, which is reasonably close to the value of 0.75 nm used here. In a similar way, values reported for the interbase distance b_0 vary considerably, since they depend on the precise DNA sequence under study and the physical conditions of the solution in which the DNA is immersed. While an inter-phosphorus distance of 0.59 nm has been established for ssDNA⁵⁶, stacking of bases leads to an interbase distance that is on average shorter^{57,58}; motivated by the findings of Ref. 55, we choose a value of $b_0 = 0.43$ nm. Therefore, the contour length of the ssDNA strands used in the present study is $l_{\text{contour}} \sim 27.5$ nm. With both l_{contour} and $l_{\text{Kuhn}} = 2p_{\text{ss}} = 1.5$ nm ready at hand, we find that $n_K = 18$, $n_{\text{se}} = 4$, and consequently $n_{\text{sp}} = 14$. For an isolated DNA strand, we find a radius of gyration of $R_g \sim 0.7R_C$, thus the radius of gyration of the DNA strands is indeed of the order of the colloidal size.

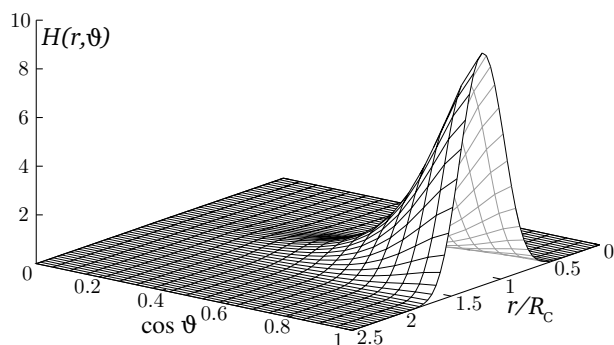


Fig. 2 The height distribution $H(r, \vartheta)$ of reactive ends, as function of (a) the distance r of the centre of mass of the last $2n_{\text{se}}$ Kuhn segments of a strand, \mathbf{r}_{end} , from its anchoring point on the colloid, $\mathbf{r}_{\text{anchor}}$; and (b) the deviation in angle ϑ between the vectors $\mathbf{r}_{\text{end}} - \mathbf{r}_{\text{anchor}}$ and the connection vector from the colloid's centre to the anchoring point. This distribution is peaked at ($r \sim 1.2R_C$, $\vartheta = 0$).

The effective line charge density ν is interpolated from Table 1 in Ref. 50 and we find $\nu(I = 0.21 \text{ mol/l; ssDNA}) = 2.07 \text{ e/nm}$, translating to a charge of $q \sim 3 \text{ e}$ per vertex. In our study, we use $D = 80$ for all temperatures and the Debye screening length varies from 0.67 nm (at 25 °C) to 0.71 nm (at 65 °C).

2.3 Simulations

To study the behaviour of an isolated DNACC, we implement Monte Carlo simulations utilising crankshaft and pivot moves to equilibrate the ssDNA chains. Every few Monte Carlo sweeps, we also try to regrow whole chains by employing configurational bias Monte Carlo simulations⁵⁹.

These simulations then allow us to gain insight into the height distribution $H(r, \vartheta)$ of the ends of the DNA strands. For convenience, we measure this distribution as a function of two parameters: (a) the distance r of the centre of mass of the last $2n_{\text{se}}$ Kuhn segments of a strand, \mathbf{r}_{end} , from its anchoring point on the colloid, $\mathbf{r}_{\text{anchor}}$. The choice of $2n_{\text{se}}$ will be motivated in Sec. 3; (b) the deviation in angle ϑ between the vectors $\mathbf{r}_{\text{end}} - \mathbf{r}_{\text{anchor}}$ and the connection vector from the colloid's centre to the anchoring point. This distribution captures how sticky ends are restricted in their movement due to the fact that the ssDNA strands are tethered to the colloid and due to neighbouring DNA strands, showing a peak at ($r \sim 1.2R_C$, $\vartheta = 0$) (Fig. 2).

In addition, we use a modification of Widom's particle insertion (mWPI) technique⁶⁰ to determine the steric repulsion $\Phi_{2,\text{rep}}(r, T)$ between two DNACCs separated by distance r and at temperature T in the zero density limit. We find $\beta\Phi_{2,\text{rep}}$ to be temperature-independent over a wide range of temperatures from $T = 25 \text{ °C}$ to 75 °C (Fig. 3, solid line).

The large number of degrees of freedom with which

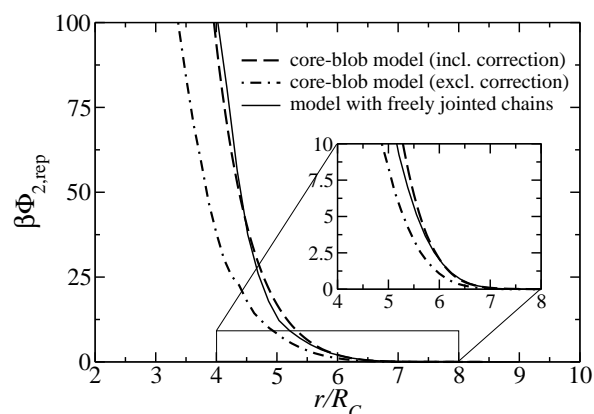


Fig. 3 The steric repulsion between two DNACCs, $\beta\Phi_{2,\text{rep}}$, as function of the distance between them and as obtained by the model with explicit DNA chains (solid line) and the core-blob model (without correction: dash-dotted line, with correction: dashed line). The repulsion is found to be independent of temperature. The inset shows a close-up of the functions for distances $r/R_C \gtrsim 5$.

DNACCs are described in the present model would render large-scale simulations of crystals of DNACCs unfeasible. Also, the present model would call for a binding scheme of the sticky ends, where several Kuhn segments align to form the dsDNA stretch^{26,31,34–36}, which is computationally rather expensive. We therefore refrain from implementing binding between complementary ssDNA sequences in the present model and rather develop a more coarse-grained model—which we term core-blob model—in the next section.

3 Stage 2: Core-blob model

The aim of the core-blob model is to arrive at the simplest possible model of DNACCs which preserves (i) the steric repulsion $\Phi_{2,\text{rep}}$ between two isolated DNACCs and (ii) the height profile of sticky ends with respect to their colloid as obtained from the model of explicit DNA chains (see Sec. 2). We therefore model each of the N_{str} sticky ends as an entity called “blob”, which is grafted to the surface of the colloid at fixed anchoring points. In this, the last $2n_{\text{se}}$ segments of every freely jointed, charged chain constitute a blob; then—to a first approximation—the blob's centre represents the connection point between the sticky end and its spacer. The gold colloid and the remaining $n_K - 2n_{\text{se}}$ segments of all N_{str} strands form the “core”, leading to a model of $N_{\text{str}} + 1$ separate entities. The model is defined by four different interactions (Fig. 4) which we derive from the model of explicit DNA chains via Monte Carlo simulations using mWPI⁶⁰ and biased simulations⁵⁹:

- (i) the repulsive interaction $\Phi_{bb}(r)$ acting between two blobs tethered to *different* colloids and a distance r apart is approximated as the interaction of two isolated (=non-tethered) freely jointed, charged chains of length $2n_{se}$ ⁶¹. As anticipated from studies of polymers, this potential is of Gaussian shape (see e.g.⁶¹). However, due to the finite length and the charge carried by the ssDNA, the repulsion found here is considerably stronger than the $2k_B T$ characteristic of polymers in the scaling regime (see Fig. 4a);
- (ii) the interaction $\Phi_{cb}(r)$ of a blob with the core of *another* colloid separated by distance r . We approximate this potential by simulating a single freely jointed, charged chain of $2n_{se}$ segments interacting with a bare core, i.e. a hard sphere grafted with chains of length $n_K - 2n_{se}$ (see Fig. 4b);
- (iii) the interaction $\Phi_{cc}(r)$ between two cores at distance r is estimated as the zero density repulsion between two colloids each grafted with chains of length $n_K - 2n_{se}$ (see Fig. 4c);
- (iv) the interaction of a single sticky end with *all* the remains of its *own* colloid (i.e. the core and all other blobs) cannot trivially be split into a repulsive and a tethering contribution due to intricate multi-body contributions. In an isolated DNACC, we therefore determine the full multi-body potential $\Phi_{scb}(r, \vartheta)$ from the height distribution of sticky ends in the underlying model as $\Phi_{scb}(r, \vartheta) = -\log H(r, \vartheta)$ (see Fig. 4d). By construction, this potential guarantees the preservation of the height profile of sticky ends with respect to the model of explicit DNA chains in the regime of dilute solutions of DNACCs (see Sec. 2).

To evaluate the reliability of the core-blob model, we determine the steric repulsion $\Phi_{2,rep}(r)$ between two isolated DNACCs within this model via mWPI⁶⁰ and compare the results to our findings from the model of explicit DNA chains. We find that the core-blob model underestimates $\Phi_{2,rep}(r)$ (see Fig. 3, dash-dotted line). The reason for this discrepancy can be traced back ‘to the fact that the core-core repulsion Φ_{cc} is too soft (Fig. 4c, dashed line), since this potential should also include a multi-body contribution from the sticky ends tethered to the spacer chains, which cannot be captured by Φ_{bb} alone. We therefore introduce a correction to Φ_{cc} (Fig. 4c, solid line) by simulating the repulsion between two colloids dressed with chains of length $n_K - n_{se}$ instead of $n_K - 2n_{se}$. Then, the core-blob model recovers $\Phi_{2,rep}$ with sufficient accuracy for all distances. Especially, we find good agreement for distances $r > 5.5R_C$ (see Fig. 3, dashed line) which includes the range of experimentally observed next neighbour distances, $r = a\sqrt{3}/2 \gtrsim 6.1R_C$, where the measured CsCl lattice spacings $a \gtrsim 42.5 \text{ nm}$ ⁷. Defining the colloidal packing

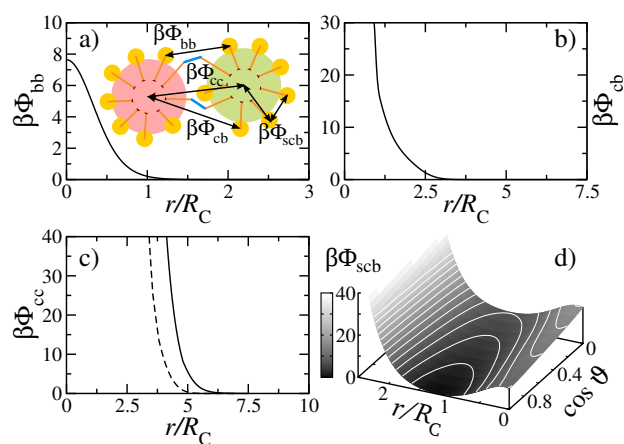


Fig. 4 (Colour online) The core-blob model, in which sticky ends are modelled as blobs and the spacers and colloids constitute a core [see inset in a)]. a) Blobs on different colloids interact via $\beta\Phi_{bb}$; b) a blob with the core of another colloid via $\beta\Phi_{cb}$; c) different cores via $\beta\Phi_{cc}$ (dashed line: uncorrected potential; solid line: corrected potential); and d) blobs with their own colloid via $\beta\Phi_{scb}$ (Isolines shown every $2.5 k_B T$). Further, blobs on colloids of different identity can bind and form short stretches of dsDNA [sketched as blue rods in the inset in a)].

fraction $\eta = \frac{4\pi}{3} R_C^3 N/V$ (with V the volume of the conventional unit cell and N the number of DNACCs in this unit cell) and assuming the experimentally observed CsCl structure to be the thermodynamically stable structure, we can determine the packing fraction up to which our model shows high reliability as $\eta = \frac{4\pi}{3} R_C^3 2/a^3$. Using $a = 25.5/\sqrt{3}R_C$, we find $\eta \lesssim 0.033$.

Finally, the hybridisation of complimentary sticky ends has to be modelled via a suitable Monte Carlo move. In this, we have to account for binding of initially unbound sticky ends, breakage of initially bound sticky ends, as well as for the change of binding partner for an already hybridised sticky end. Further, we wish to use the experimentally measured data on the DNA hybridisation free energy. Since the persistence length of dsDNA far exceeds that of ssDNA, we model the hybridised sticky ends as a (volumeless) rigid rod of fixed length $\mathcal{L} = n_{se}/l_{Kuhn}$. For simplicity, we ignore the change in inter-base distance between ssDNA (0.43 nm) and dsDNA (0.34 nm). Binding is possible between a chosen (bound or unbound) blob i on a given colloid and all unbound blobs j tethered to unlike colloids and within a distance of approach $r_{ij} < \mathcal{L}$. Upon binding, the reaction partner j is moved to a distance \mathcal{L} from i along the connection line \mathbf{r}_{ij} . Since bound blobs cannot move independently anymore, binding leads to the loss of a degree of freedom which is reintroduced upon unbinding by placing j along \mathbf{r}_{ij} with a probability of r_{ij}^2 , thereby guaranteeing detailed balance. The probabilities for each pos-

T [°C]	G_{hyb} [$k_B T$]	T [°C]	G_{hyb} [$k_B T$]
50	-14.12	61.4	-9.03
55	-11.83	62.1	-8.71
55.8	-11.47	63.2	-8.23
56.9	-10.98	64.3	-7.75
58.0	-10.49	65.1	-7.44
59.1	-10.00	66.2	-6.97
60.2	-9.51	66.9	-6.67

Table 1 The values of the hybridisation free energy G_{hyb} [$k_B T$] of the DNA strands used in system V in Ref. 7 at different temperatures (according to `DINAMelt`⁶²): the DNA spacer sequence is given as 5'-TTT TTT TTT TTT TTT TTT TTT TTT CGT TGG CTG GAT AGC TGT GTT CT-3'. Sticky ends on *A*-colloids read 5'-TAA CCT AAC CTT CAT-3', while on *B*-colloids the complementary sequence is found, 5'-ATG AAG GTT AGG TTA-3'. Values were determined at a ionic strength of 0.21 mol/l.

sible bound state ij and the unbound state i to occur are determined by their respective weights W :

$$W_{ij} = \frac{K}{\frac{4\pi}{3} \mathcal{L}^3 \rho_0} \exp(-\beta U_{ij}), \quad (5)$$

and

$$W_i = \exp(-\beta U_i), \quad (6)$$

where ρ_0 is the standard density of 1 mol/l, U_{ij} is the potential energy of the state where i is bound to j and U_i the potential energy of the state where i is unbound; K is the equilibrium binding constant of two sticky ends and is connected to the hybridisation free energy G_{hyb} via $K = \exp(-\beta G_{\text{hyb}})$; G_{hyb} (and thereby K) depends on the nucleotide sequence of the sticky ends and it is temperature- and salt-dependent; it can be approximated as the hybridisation free energy of two sticky ends free in solution (e.g. via `DINAMelt`⁶²). The values for G_{hyb} used here are given in Tab. 1. The present binding move is justified in more detail in the Appendix A.

4 Stage 3: Effective interactions

We use the core-blob model to calculate the pair interactions between two DNACCs in the zero density limit. Interactions between like colloids (i.e. *AA* and *BB*) are described by the purely repulsive, temperature-independent steric repulsion between two DNACCs, $\beta\Phi_{2,\text{rep}}$, calculated above. The interaction of unlike colloids (i.e. *AB*) additionally features an attractive potential $\Phi_{2,\text{hyb}}$ stemming from hybridisation of DNA strands. Thus,

$$\beta\Phi_2^{AA/BB}(r) = \beta\Phi_{2,\text{rep}}(r) \quad (7)$$

and

$$\beta\Phi_2^{AB}(r, T) = \beta\Phi_{2,\text{rep}}(r) + \beta\Phi_{2,\text{hyb}}(r, T). \quad (8)$$

$\beta\Phi_{2,\text{hyb}}$ can be obtained by evaluating

$$\beta\Phi_{2,\text{hyb}}(r, T) = - \int_{\beta G_{\text{hyb}}}^{\infty} d(\beta G'_{\text{hyb}}) \langle \zeta(r) \rangle_{K=\exp(-\beta G'_{\text{hyb}})}, \quad (9)$$

as described in Ref. 32. Here, $\zeta(r)$ is the number of DNA bridges formed between the two colloids at fixed distance r , and $\langle \dots \rangle$ denotes the statistical average. In practise, the integration is performed between G_{hyb} of interest and a value sufficiently large for sticky ends not to hybridise anymore.

Due to the temperature-dependence of G_{hyb} , the depth of the minimum of $\Phi_2^{AB}(r, T)$ varies strongly with temperature. For instance, a change in temperature from 62.1 to 56.9 °C results in a drop in the minimum of Φ_2^{AB} of roughly 20 $k_B T$ (corresponding to 13 kcal/mol). This strong temperature dependence of the DNA-mediated attraction explains the difficulty in crystallising the DNACCs: only in a narrow temperature range around $\sim 65^\circ\text{C}$ is the minimum in $\Phi_2^{AB}(r, T)$ shallow enough to allow for the formation *and* breakage of DNA links. Upon lowering T , the bonds that form cannot break anymore and the system gets stuck in disordered aggregates, even if an ordered structure is thermodynamically stable¹.

To assess the predictive power of the core-blob model, it would be desirable to compare $\Phi_2^{AA/BB}$ and Φ_2^{AB} to experimental results. But while such potentials can be experimentally determined for *micron*-sized colloids by using optical tweezers^{12,43}, the same is not feasible for nano-colloids. Other experimental validation techniques will be required. Thus, validation of our model against experimental data will only be studied later, by comparing the computed and experimentally determined phase diagrams.

Thus far, our approach has allowed us to compute the effective pair potential between DNACCs. Pairwise additive interactions are typically used to model DNACCs as structureless particles in theoretical studies (e.g. Ref. 22) and also in some computational studies (e.g. Ref. 33). However, in the regime where $R_g \sim R_C$ we may expect that the assumption of pairwise additivity breaks down. With the present model we can quantify the importance of such many-body interactions.

5 Three-body interactions

Using the core-blob model we test if the three-body interactions of a system of two *A* colloids and one *B* colloid can be written as the sum of the various two-body contributions. It can be anticipated that differences between the three-body interaction and the sum of the two-body contributions will mainly arise from a competition of the two *A* colloids for the sticky ends of the *B* colloid and will therefore crucially depend on the arrangement of the DNACCs with respect to each

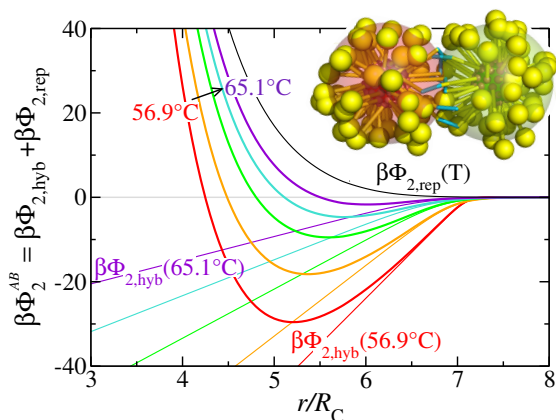


Fig. 5 (Colour online) The effective interaction between unlike colloids $\beta\Phi_{2,\text{rep}}^{AB}$ (bold lines) is the sum of the temperature-independent steric repulsion $\beta\Phi_{2,\text{rep}} (= \beta\Phi_{2,\text{rep}}^{AA/BB})$ between two colloids, and a temperature-dependent attractive potential $\beta\Phi_{2,\text{hyb}}$ arising from hybridisation of DNA strands. Data are shown for different temperatures (top to bottom: 65.1°C, 63.2°C, 61.4°C, 59.1°C, and 56.9°C). The inset shows a simulation snapshot of an A colloid (green) interacting with a B colloid (red). Unhybridised sticky ends are shown as yellow spheres, while DNA bridges are shown as blue rods. The translucent spheres indicate the average position of the sticky ends.

other. A linear arrangement of the colloids, with the B colloid positioned between the two A colloids, is expected to lead to little discrepancy since in the regime studied here ($R_g \sim R_C$), the DNA strands are too short to reach complementary sticky ends at the back of the DNACC they face.

However, in the typical arrangements of DNACCs that occur in a crystalline environment, many-body effects are more likely to arise. In a crystal, a given colloid is typically surrounded by several DNACCs of the other species that all compete for the strands of the central colloid. It is therefore interesting to study the three-body interaction $\Phi_3^{ABA} = [\Phi_3^{BAB}]$ for three DNACCs arranged in an equilateral triangle of side-length r and test if the following relation holds

$$\beta\Phi_3^{ABA}(r, T) \stackrel{?}{=} 3\beta\Phi_{2,\text{rep}}(r) + 2\beta\Phi_{2,\text{hyb}}(r, T). \quad (10)$$

5.1 Methods

In analogy to the AB two-body interactions, also the ABA/BAB three-body interaction can be split into a contribution $\Phi_{3,\text{rep}}$ stemming from steric repulsions between the DNACCs, and an attractive part $\Phi_{3,\text{hyb}}$ arising from DNA strand hybridisation, i.e.

$$\beta\Phi_3^{ABA}(r, T) = \beta\Phi_{3,\text{rep}}(r, T) + \beta\Phi_{3,\text{hyb}}(r, T). \quad (11)$$

We first calculate $\Phi_{3,\text{rep}}$ by generalising the mWPI⁶⁰: we place three *non-reactive* DNACCs, i.e. DNACCs with non-binding sticky ends, in an equilateral triangle of side-length $r = r_{\text{max}}$ bigger than the expected range of the interactions. Then we repeatedly reduce the side-length to $r - \Delta r$ and measure $\langle \exp[-\beta\Delta U(r \rightarrow r - \Delta r)] \rangle$, where $\Delta U(r \rightarrow r - \Delta r)$ is the change in potential energy between the three DNACCs due to the move. In this way, we move the DNACCs together. Then, $\beta\Phi_{3,\text{rep}}(r) = -\sum_{r'=r_{\text{max}}}^{r+\Delta r} \log \langle \exp[-\beta\Delta U(r' \rightarrow r' - \Delta r)] \rangle$.

To determine $\Phi_{3,\text{hyb}}(r)$, we arrange the three DNACCs in an equilateral triangle of fixed side-length r . Measuring the total number of DNA bridges formed at this distance, $\zeta(r)$, we can calculate $\Phi_{3,\text{hyb}}(r)$ as

$$\beta\Phi_{3,\text{hyb}}(r) = -\int_{\beta G_{\text{hyb}}}^{\infty} d(\beta G'_{\text{hyb}}) \langle \zeta(r) \rangle_{K=\exp(-\beta G'_{\text{hyb}})}, \quad (12)$$

in analogy to the determination of $\Phi_{2,\text{hyb}}$ (see Eqn. 9).

5.2 Results

As can be seen from Fig. 6, the repulsive three-body potential $\beta\Phi_{3,\text{rep}}(r)$ is, to a good approximation, equal to the sum of the two-body contributions $3\beta\Phi_{2,\text{rep}}(r)$. However, pairwise additivity does not hold for the attractive part of the three-body potential, $\beta\Phi_{3,\text{hyb}}$. As anticipated, the two A colloids increasingly compete for the available sticky ends of B as the colloids are moved closer together and fewer bonds can form for each of the two AB pairs than in an isolated, single AB pair. This overestimation of formed DNA bridges within the pair potential approach directly translates to an overestimate of the depth of the attraction between the colloids (cf. Eqn. 9 and see Fig. 6) and consequently an underestimate of the position of the minimum in the attraction. We therefore expect that an analysis of DNACC crystals within the pair potential framework will predict more compact crystals than found experimentally⁷.

6 Phase behaviour of DNA coated colloids

To assess the predictive power of both the core-blob model and the pair potential approach, we study the phase behaviour of DNACCs by implementing free-energy calculations within both approaches. The results are then compared to data available from experiments⁷, such as the stable crystal structure, its lattice constant and the melting temperature of these crystals.

6.1 Crystal structure prediction

As was already known to Ostwald⁶³, observing spontaneous formation of a crystal does not imply that the observed structure has the lowest free energy. Rather, we have to consider the thermodynamic stability of all possible crystal struc-

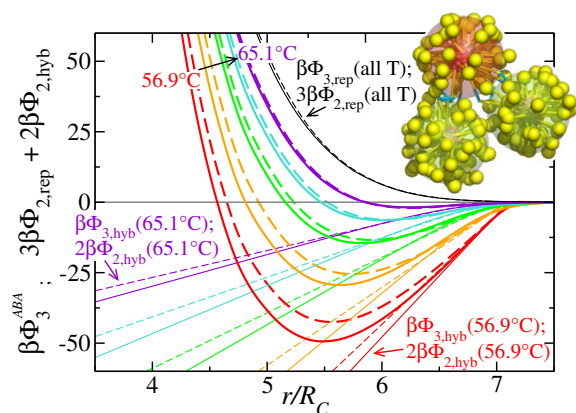


Fig. 6 (Colour online) The three-body effective interaction $\beta\Phi_3(r)$ (bold dashed lines) of two *A* and one *B* colloids arranged in an equilateral triangle compared with the sum of the two-body contributions $3\beta\Phi_{2,rep} + 2\beta\Phi_{2,hyb}$ (bold solid lines) at 65.19°C, 63.29°C, 61.49°C, 59.1°C, and 56.9°C (top to bottom). Further, the repulsive contribution and the attractive hybridisation contributions to the effective interactions are shown explicitly in thinner lines (two-body contributions: solid lines, three-body contributions: dashed lines). The inset shows a simulation snapshot. *A* colloids are shown as green spheres, and the *B* colloid as a red sphere. Unhybridised sticky ends are shown as yellow spheres, while DNA bridges are sketched as blue rods. The translucent spheres indicate the average position of the sticky ends.

tures. To identify credible candidates for the most stable crystal structure, we use optimisation techniques based on genetic algorithms^{64,65}. The structures that the genetic algorithm identifies as plausible are then considered in the free-energy calculations. We adapt a search strategy for 2D binary mixtures^{66,67} to 3D, augmenting it with a parametrisation of search space that excludes *a priori* configurations with overlapping colloids⁶⁸. The lattice parameters describing the crystal structures are encoded in binary individuals and a random crossover is employed as mating scheme. Mutations take place with a rate of 0.05. We limit our search to symmetric *AB*-mixtures and lattice structures with up to eight particles per unit cell. Particles interact via their respective pair interactions (see Sec. 3), which we first fit to analytical functions (see Appendix B). Calculations are run at constant pressure P , so that the volume fraction η enters the optimisation as an independent parameter. Structures are optimised with respect to the Gibbs free energy of the system $G = U + PV - TS$, with U the internal energy, V the volume of the system, and S the entropy. In the genetic algorithm, entropy is neglected, and hence the Gibbs free energy is equal to the enthalpy of the structures. Temperature-dependence of the system only enters our approach via the temperature-dependence of the pair potential; this limitation in treating entropic effects necessitates the subsequent free energy calculations (see Sec. 6.2). To determine the minimum enthalpy configuration at a given pressure we evaluate 1000 generations of a population of 50 individual ordered structures each. Details on the general working principles of the method can be found in Ref. 65.

The genetic algorithm calculations predict the CsCl (B2) structure as the most stable one for low colloidal packing fractions η , while it also predicts a competing NaTl (B32) structure for higher η . The latter structure is of general interest since it is composed of two interpenetrating diamond structures. If it were possible to remove one of the two colloidal species in a post-assembly modification step, a diamond structure could be created¹⁵. Such assembly strategies are explored by e.g. substituting the gold colloids in one of the two species of DNACCs by organic compounds¹⁶. However, in the present system an experimental distinction between CsCl and NaTl structures would prove challenging: here, *A* and *B* colloids only differ by their sticky ends while X-ray scattering only detects the gold colloids. As a result, both CsCl and NaTl structures would experimentally be detected as bcc arrangements.

Apart from the CsCl and NaTl structures, we chose to consider a few more candidate structures: CuAu (L1₀), NaCl (B1), ‘straight’ hcp (s-hcp)³³, ZnS (B3; diamond)²², AuCd (B19), as well as substitutionally disordered CsCl and CuAu crystals^{6,34} (Fig. 7).

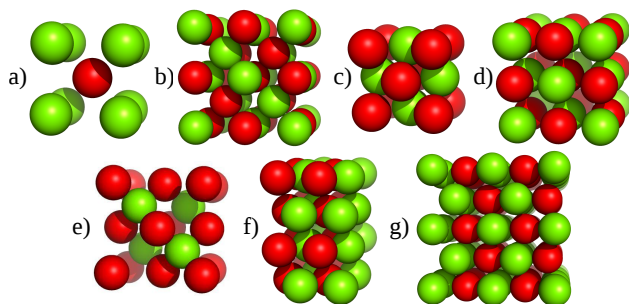


Fig. 7 (Colour online) The different crystal structures considered in this study. a) CsCl; b) NaTl; c) CuAu; d) NaCl; e) ZnS; f) 'straight' hcp; and g) AuCd. Different colloidal species are coloured red and green; for clarity, the DNA strands have been omitted.

6.2 Free energy simulations

To study the phase behaviour of DNACCs and to determine the stable crystal structure, we have to determine the free energies F of all candidate structures mentioned above. Within the effective pair potential framework, we calculate F via thermodynamic integration in the canonical ensemble as detailed in^{69,70} and using systems of at least $N = 1000$ DNACCs. Within the core-blob model, the thermodynamic integration is achieved in two steps similar to the approach followed in Ref. 30, using systems of more than $N = 100$ DNACCs. In the first step of the integration we arrange N DNACCs in the desired crystal structure at fixed colloidal volume fraction η and temperature T (and thereby fixed binding free energy G_{hyb} ; see Tab. 1). We then transform this DNACC crystal into a system of non-binding DNACCs artificially fixed to their lattice sites. We achieve this by gradually increasing G_{hyb} so that sticky ends do not bind anymore; at the same time, we gradually confine the centres of the colloids to individual, small cells of volume v around the perfect lattice sites \mathbf{R}^{LS} of the chosen crystal structure by raising a potential barrier. For a colloid i centred at \mathbf{R}_i^C and at integration point $\lambda \in [0, 1]$, the potential energy $\phi_{\text{barr}}^\lambda$ due to the barrier is given as

$$\beta\phi_{\text{barr}}^\lambda(\mathbf{R}_i^C) = \begin{cases} 0 & \text{if } \mathbf{R}_i^C \in \cup_j v(\mathbf{R}_j^{\text{LS}}) \\ \lambda\beta U_{\text{barr}} & \text{else,} \end{cases} \quad (13)$$

where $\cup_j v(\mathbf{R}_j^{\text{LS}})$ is the union of the confining volumes around all perfect lattice sites \mathbf{R}^{LS} and U_{barr} is the maximal height of the barrier. U_{barr} is chosen sufficiently high for the crystals not to melt during the thermodynamic integration. The total potential energy of the system due to the barrier is given as $\beta\Phi_{\text{barr}}^\lambda = \sum_{i=1}^N \beta\phi_{\text{barr}}^\lambda(\mathbf{R}_i^C)$. Via Gauss-Lobatto quadrature⁷¹,

we numerically evaluate

$$\beta\Delta F_1 = \int_1^0 d\lambda \langle \beta\Phi_{\text{barr}}^\lambda \rangle - \int_{\beta G_{\text{hyb}}}^\infty d(\beta G')_{\text{hyb}} \langle \zeta \rangle_{K=\exp(-\beta G'_{\text{hyb}})}, \quad (14)$$

where ΔF_1 is the difference in free energy between the crystal of interest and the crystal of inert, confined DNACCs. In practise, the upper limit of the second integral is replaced by a hybridisation free energy sufficiently large to guarantee that no hybridisation of the ssNDA sticky ends takes place. ζ denotes the total number of DNA bridges formed in the system.

In the second integration step, we use the lattice-coupling expansion method⁷²: we linearly expand both the crystal of inert DNACCs and the potential barrier—which guarantees to hold the crystal in place—i.e.

$$\mathbf{R}_i^X \rightarrow \gamma \mathbf{R}_i^X \quad (15)$$

with $X = C, \text{LS}$ and the expansion factor $\gamma \in [1, \infty]$. In practise, the infinite expansion is approximated by expanding the system sufficiently for particles not to interact anymore. The free energy difference between the unexpanded and the expanded crystal of inert DNACCs is given by

$$\beta\Delta F_2 = \beta \int_1^\infty d\gamma \langle \mathcal{W} \rangle_\gamma, \quad (16)$$

where \mathcal{W} is a modified virial given as

$$\mathcal{W} = \sum_{i < j} \mathbf{f}_{ij,\gamma} \mathbf{R}_{ij,\eta}^C. \quad (17)$$

Here, $\mathbf{f}_{ij,\gamma}$ is the force acting between DNACC i and DNACC j at expansion factor γ , and $\mathbf{R}_{ij,\eta}^C = \mathbf{R}_{j,\eta}^C - \mathbf{R}_{i,\eta}^C$ is the separation of the two DNACCs at the *original* colloidal packing fraction η (i.e. at $\gamma = 1$)⁷².

The total free energy per colloid is then given as

$$\beta F/N = \beta f_{\text{id}} + \beta\Delta F_1/N + \beta\Delta F_2/N, \quad (18)$$

where βf_{id} is the free energy of an isolated DNACC; being the same for all crystal structures, this contribution can be neglected in the determination of the thermodynamically stable crystal structure.

6.3 Results

For all candidate crystal structures, we determine the free energy for a range of packing fractions, concentrating on the regime where $T \lesssim 65^\circ\text{C}$. This is the temperature where the effective interaction Φ_2 develops a minimum indicating that at these temperatures DNA bridges form between DNACCs.

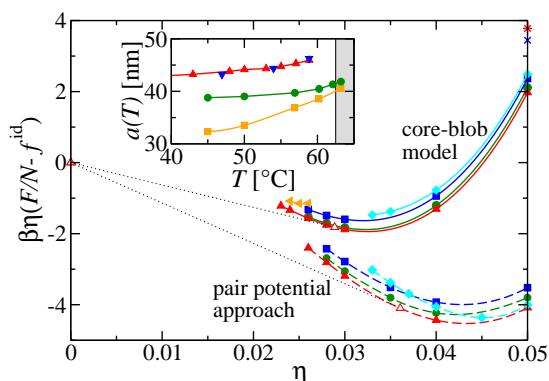


Fig. 8 (Colour online) The excess free energy $\beta\eta(F/N - f^{\text{id}})$ as function of the colloidal volume fraction η at $T = 56.9^\circ\text{C}$. Both the core-blob model (solid lines) and the pair potential approach (dashed lines) find the CsCl structure (▲) to be the most stable one. This crystal is in equilibrium with a dilute vapour (△). The common tangents are shown as dotted lines. In contrast, the liquid (◄) is found to be only metastable. In close competition with the CsCl structure, a range of metastable crystal structures is found: s-hcp: ●, CuAu: ■, NaTl: ◆, disordered CsCl: *, disordered CuAu: ×. Inset: A comparison of the lattice constant a (in nm) as function of temperature T (in $^\circ\text{C}$) as measured in the experiments (heating: ▲, cooling: ▼⁷) and as obtained from simulations with the core-blob model (●) and using the pair potential approach (■). The shaded region indicates the temperatures at which crystals are not stable in experiments.

In this regime, crystals of low packing fraction are expected to be stabilised by DNA hybridisation.

We first present results obtained via the core-blob model (Fig. 8, solid lines), initially concentrating on a temperature of $T = 56.9^\circ\text{C}$. We find that at all packing fractions η considered, the CsCl structure is the most stable, in agreement with experimental findings⁷. In close competition with the CsCl structure, we find a whole range of metastable crystal structures, namely s-hcp, CuAu and NaTl. These structures are mechanically stable for packing fractions of $\eta \gtrsim 0.026 - 0.035$; by contrast, the CsCl structure is already found stable for $\eta \gtrsim 0.023$. At high $\eta \sim 0.065$, a metastable AuCd phase appears. By contrast, both the NaCl and ZnS structures are found to be mechanically unstable and melt at all volume fractions considered. Extrapolation of the collected data suggests that the NaTl structure would become the stable structure for $\eta \gtrsim 0.07$. The reason for this transition from CsCl to NaTl can be understood from studying the mechanism stabilising the latter structure: we found that NaTl crystals are mechanically stable only at packing fractions $\eta \gtrsim 0.035$, where colloids not only form DNA bridges with next neighbour colloids, but also with the unlike colloids found in the second coordination shell. By contrast, CsCl cannot form DNA bridges with the next nearest neighbours since they are all like colloids. At sufficiently high packing fractions beyond $\eta \sim 0.05$, the excluded volume effects of the DNA strands render binding to the nearest neighbours increasingly challenging for both CsCl and NaTl. While this severely limits the amount of possible DNA bridges for CsCl, NaTl can instead bind to the next nearest neighbours, allowing it to eventually become the thermodynamically stable structure around $\eta \gtrsim 0.07$. However, such high volume fractions cannot be achieved in experiments where crystals form from a dilute vapour. Moreover, we stress that the extrapolation should be taken with a grain of salt as we do not expect our model to be fully valid at these high densities (see Sec. 3).

Next, we study substitutionally disordered crystals. At low volume fractions $\eta < 0.05$, such crystals have only limited mechanical stability and few substitutional changes can be sustained by the crystals. A single substitutional defect, in which only a single pair of neighbouring A and B colloids exchange sites in an otherwise perfect crystal increases the free energy of the crystal by $\Delta F = 0.5k_B T$ ($\eta = 0.024$) to $1.6k_B T$ ($\eta = 0.05$) for CsCl and by $\Delta F = 0.1k_B T$ ($\eta = 0.026$) to $1.0k_B T$ ($\eta = 0.05$) for CuAu. At higher $\eta \sim 0.05$, more substitutional disorder can be stabilised: however, the free energy of the substitutionally disordered structures is higher than that of the perfectly ordered crystals. This can be understood by considering the contributions to the free energy. While $\beta\Delta F_2$ (Eqn. 16) and the first integral of $\beta\Delta F_1$ (Eqn. 14) are independent of the distribution of the A and B colloids over lattice sites, the second integral determining $\beta\Delta F_1$ (Eqn. 14) depends crucially on the average number of DNA bridges $\langle \zeta \rangle$ formed

in the crystal. As $\langle \zeta \rangle$ is lower in a substitutionally disordered crystal than in a perfectly ordered crystal, $\beta \Delta F_1$ is lower for the latter systems. Still, substitutionally disordered crystals have been seen to form spontaneously in simulations³⁴. The present results suggest that these structures are kinetically arrested for $T \lesssim 65^\circ\text{C}$.

At high $\eta > 0.1$, it is to be anticipated that the phase behaviour of DNACCs will increasingly be dominated by the steric repulsion from the compressed DNA strands as well as excluded volume interactions of colloids. Then, close-packed structures such as CuAu or s-hcp are expected to be thermodynamically stable. Since crystals are not stabilised by DNA hybridisation anymore in this regime, substitutionally disordered crystals should be favoured over ordered structures. We emphasise that exploration of this regime is beyond the scope of the current contribution.

Compared to the results of the core-blob model, the pair potential approach—as expected—underestimates the free energy by overestimating the number of DNA bridges formed in the system (Fig. 8, dashed lines). Still, it offers a good estimate of the range of mechanical stability of the various crystal structures and predicts the same phase order as the core-blob model, i.e. CsCl as most stable structure, followed by metastable s-hcp, CuAu and NaTl. However, within the pair potential approach, the NaTl structure is found to out-compete the CsCl structure already around lower packing fractions of $\eta \gtrsim 0.055$. Structures found to be mechanically unstable in the core-blob approach (ZnS, CuAu) are also found unstable within the current framework. Therefore, the pair potential approach offers an excellent tool for assessing the mechanical stability of the candidate structures and can be used as a pre-selection tool for choosing the crystal structures to be studied with the core-blob approach.

The next question is: what happens when crystals of DNACCs melt? Do they form a dilute vapour or a dense liquid? To answer this question, we determined the melting behaviour of the CsCl structure for both models, and for both we find that the crystal coexists with a dilute vapour. A similar conclusion was reached in Ref. 30. We find no evidence for a phase transition between the dilute solution and a denser liquid phase. Using the common-tangent construction (see Fig. 8), we can determine the colloidal volume fraction of the CsCl structure at coexistence, η_s . Within the core-blob approach, we find $\eta_s = 0.029$ at $T = 56.9^\circ\text{C}$. As expected, the pair potential approach predicts a more compact equilibrium CsCl crystal of $\eta_s = 0.036$. Determining η_s within both models for several temperatures, we can then determine the prediction of the lattice constant a of the equilibrium CsCl crystals as a function of temperature

$$a(T) = \sqrt[3]{\frac{8\pi}{3\eta_s(T)}} R_C. \quad (19)$$

The computed values of $a(T)$ are compared to experimental data in the inset of Fig. 8. Concentrating first on the core-blob approach, we find that it predicts the thermal expansion coefficient of the crystals at least qualitatively correctly. It also correctly predicts that at sufficiently low temperatures a levels off to a constant value. However, the simulations predict denser crystals than experimentally observed. The discrepancy in lattice constant is as large as $\sim 12\%$. This observation points to a problem in the input in our model. One drawback is that we neglected the hexane-thiol linker grafting the DNA strands to the colloids, which has an end-to-end length of ~ 0.8 nm. Taking this linker into account is expected to reduce the discrepancy to the experiments slightly. However, the major weak spots are the choices of the ‘preferred’ experimental values for the persistence length p_{ss} and for the interbase distance b_0 of ssDNA. We chose an average value for both parameters but, in reality, both numbers are expected to depend on the precise base sequence of the ssDNA^{54,58}. With more systematic experimental data on the sequence dependence of these values, we expect that the core-blob model would also allow for quantitative predictions of spatial quantities (such as the lattice constant).

We note that the pair potential approach captures neither the length nor the temperature behaviour of DNACC crystals correctly: it seriously overestimates the thermal expansion coefficient of the crystals. Hence, the pair-potential approach cannot be used to describe the thermal properties of DNACC crystals.

For completeness, we point out that in the temperature regime where DNA strands cannot bind anymore (i.e. well above 65°C), crystallisation can only occur due to excluded volume effects for which a high osmotic pressure is needed. Since the focus of our work was on crystal formation triggered by DNA hybridisation, we did not study this regime.

The equilibrium melting (‘sublimation’) temperature of a DNACC crystal depends on the concentration of the dilute solution and, in contact with an infinitely dilute solution, all DNACC crystals will eventually evaporate. However, the rate at which this happens depends strongly on temperature. The more relevant question is therefore: at what temperature does the rate of sublimation of a DNACC crystal become experimentally observable? Experimentally, the effective melting temperature was determined via ultraviolet-visible spectrophotometry in Ref. 7 and found to be $T_m = 62.5(\pm 0.3)^\circ\text{C}$. To be able to compare simulational data to experiments, we can estimate the temperature below which the spontaneous evaporation of DNACCs from a crystal becomes negligible. A rough estimate of the concentration where this happens can be obtained using Smoluchowski’s treatment of the diffusion limited growth of a cluster^{73,74}. In equilibrium, the evaporation rate of DNACCs from a solid, spherical cluster of radius \hat{R} equals the aggregation rate to that cluster in the presence of

a dilute vapour of DNACCs of density ρ_v . We can determine the latter rate, dN/dt , as

$$\frac{dN}{dt} = 4\pi\rho_v\hat{D}\hat{R}, \quad (20)$$

with \hat{D} being the diffusion constant of a DNACC^{73,74}. On the other hand, the amount of DNACCs in the solid cluster can be expressed as $N = \frac{4\pi}{3}\hat{R}^3\rho_s$, where ρ_s is the density of the cluster and $\rho_s = \eta_s/(\frac{4\pi}{3}R_C^3)$. Then, the following relation between ρ_v and ρ_s can be derived:

$$\frac{\rho_v}{\rho_s} = \frac{1}{2\hat{D}} \frac{d\hat{R}^2}{dt} \quad (21)$$

To estimate ρ_v , we assume that a solid cluster grows to $1\ \mu\text{m}$ in one day ($\sim 10^5\text{s}$). Further, we estimate the diffusion constant of a DNACC from the Einstein–Smoluchowski relation, $\hat{D} = k_B T / (6\pi\hat{\eta}R_C)$. With $\hat{\eta} \sim 1\text{cP}$ being the viscosity of water, we find that $\hat{D} \sim 4 \times 10^{-7}\text{cm}^2/\text{s}$. Then, $\rho_v/\rho_s \sim 10^{-7}$. From the knowledge of the coexistence densities of the vapour and solid phases at several temperatures, we find that the last relation is fulfilled for $T_m = 63.5(\pm 0.2)\text{ }^\circ\text{C}$ in the pair potential approach and for $T_m = 61.9(\pm 0.2)\text{ }^\circ\text{C}$ in the core-blob model (see Fig. 9), which is in good agreement with the experimental finding of $T_m = 62.5(\pm 0.3)\text{ }^\circ\text{C}$. We note that in Ref. 37, we estimated the melting temperature within the core-blob approach by considering the point where small colloidal crystals melted on the time scale of a simulation. That approach is likely to lead to a higher estimate of the melting temperature and, indeed, we found that simulated crystals melt for $T \gtrsim 64.3(\pm 0.5)\text{ }^\circ\text{C}$ (see Fig. 9).

7 Conclusion

In this paper we described a method to construct a quantitative coarse-grained model of DNACCs and compared its results to experimental data. For comparison, we also studied the predictive power of a simpler model based on the effective interactions between pairs of DNACCs.

We found that the pair potential approach can be used as a qualitative tool, allowing to roughly delimit the range of mechanical stability of DNACCs, and give a coarse estimate of the melting temperature. Further, it allows for qualitative insight into the compression behaviour of DNACC crystals upon temperature reduction. This model has the advantage of being computationally inexpensive, therefore allowing for fast and large-scale testing of DNACC designs.

However, we showed that for quantitative insight into DNACC systems in the regime where the radius of gyration of the tethered DNA strands is of the order of the size of the (nano-)colloid, an explicit description of sticky ends is needed to capture the competition of DNACCs for DNA bridges. We

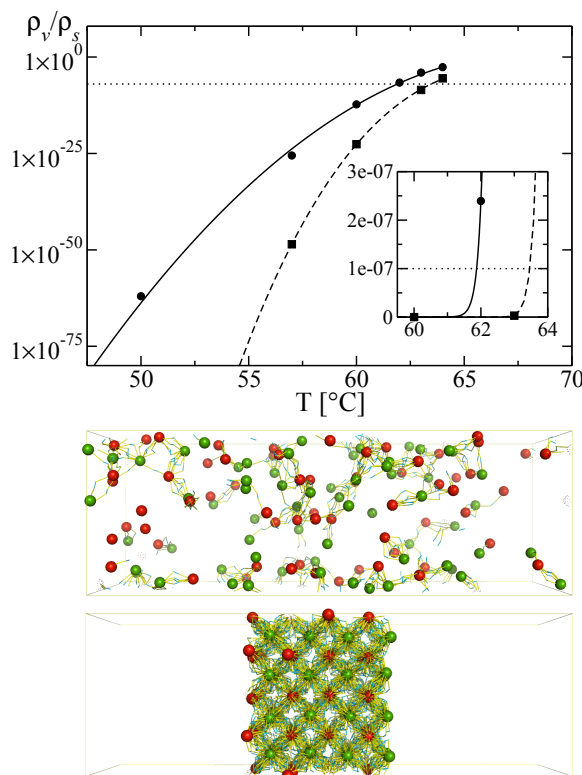


Fig. 9 Top: The ratio between the coexistence densities of the vapour and the solid, ρ_v/ρ_s , as function of the temperature according to simulations within the core-blob model (●) and the pair interaction approach (■). The solid and dashed lines serve as guides to the eye. The threshold value of 10^{-7} (dotted line) indicates the temperature below which spontaneous evaporation of DNACCs from a crystal becomes negligible. Middle and bottom: Simulation snapshots of a DNACC crystal (core-blob model) at $T = 65.1\text{ }^\circ\text{C}$ which melted during the course of the simulation and of a crystal at $T = 60.2\text{ }^\circ\text{C}$, where the evaporation of DNACCs is negligible. A colloids are shown as green spheres, B colloids in red. For clarity, only hybridised DNA chains are shown.

therefore developed a more detailed model, which we termed “core-blob model” and which is solely based on experimental input. Results from this approach show good agreement with experimental data in temperature-dependent quantities. While the lattice constant is not captured quantitatively in absolute values, the thermal expansion coefficient of crystals is well described. We speculate that, once more systematic experimental data on the persistence length and inter-base distance of ssDNA become available, the core-blob model should also account for the experimentally observed lattice spacings. We could of course have adjusted the model parameters to account for the observed lattice spacing, but this would have defeated the purpose of the present work, which was to construct a model based exclusively available ssDNA data - our model contains no *a posteriori* fitted parameters.

A potential drawback of the core-blob model is that the input parameters depend on the ssDNA length and sequence. For a given set of input parameters, new simulations are needed to redevelop the interaction potentials of the core-blob model and to predict the phase behaviour. In some cases, the situation may be better. For instance, in e.g. Ref. 8, different systems were generated by supplementing *one* choice of DNACCs with different linkers. Then, the length of spacers, their binding strength and the number of reactive ends can be tuned in a straightforward way via these linker sequences alone. Similarly, our model can be easily generalised to incorporate linkers while reusing the present representation of the DNACCs themselves. Furthermore, the core-blob model can be adapted to study e.g. systems of more complex coatings, asymmetric mixtures, or polydisperse systems, while allowing for direct mapping to the corresponding experimental system.

In summary, a rough scanning of the phase behaviour of DNACC designs via the pair potential approach can be used to preselect promising DNACC designs, which can then be quantitatively studied in more detailed, but also computationally more expensive calculations using the core-blob model. In this way, the approach presented here offers a path to computer-aided design of suitable DNA-grafted building blocks, advancing the efforts of constructing truly complex self-assembling structures.

8 Acknowledgements

We thank P. Varilly for helpful discussions and careful reading of the manuscript. Further, we thank B. Capone (Vienna), S. Angioletti-Uberti, B. M. Mognetti, W. Jacobs, G. Day (Cambridge), A. Tkachenko, D. Nykypanchuk and O. Gang (Brookhaven) for useful discussions at various stages of this project. BMM acknowledges EU funding (FP7-PEOPLE-IEF-2008 No. 236663 and FP7-PEOPLE-CIG-2011 No. 303860) and funding via the MFPL Vienna International Post-Doctoral Program for Molecular Life Sciences (funded

by Austrian Ministry of Science and Research and City of Vienna, Cultural Department - Science and Research). AD was supported by an EMBO longterm fellowship. DF and FJMV acknowledge support of ERC Advanced Grant 227758. DF acknowledges a Wolfson Merit Award of the Royal Society of London and EPSRC Programme Grant EP/I001352/1.

9 Appendix A: The binding move in the core-blob model

In the following, we derive the Monte Carlo algorithm of the binding move used for hybridisation of sticky ends in the core-blob model. In Sec. 9.1, we study the requirement of detailed balance for the case of an unbound sticky end binding to a complementary, unbound sticky end within reach, i.e. within distance of hybridised sticky ends, \mathcal{L} . Without loss of generality, we name the chosen sticky end a and assume that it is attached to a colloid of kind A . Consequently, we term its possible binding partner b , which is fixed on a colloid of kind B . We then generalise to an arbitrary number of possible binding partners (incl. the possibility of a partner change) in Sec. 9.2.

9.1 Detailed balance for one possible binding partner

To fulfil detailed balance, we need to justify that the flow \mathcal{K} from the configuration where a is unbound (“ a ”) to the state where a and b are hybridised (“ ab ”) is the same as the reverse flow, i.e.

$$\mathcal{K}(a \rightarrow ab) = \mathcal{K}(ab \rightarrow a). \quad (22)$$

We can write

$$\mathcal{K}(a \rightarrow ab) = P_a P_{\text{gen}}(a \rightarrow ab) P_{\text{acc}}(a \rightarrow ab), \quad (23)$$

where P_a is the probability of a being unbound, $P_{\text{gen}}(a \rightarrow ab)$ is the probability that the Monte Carlo move hybridises a with b and $P_{\text{acc}}(a \rightarrow ab)$ is the probability of accepting this move. An analogous formula can be written for $\mathcal{K}(ab \rightarrow a)$.

Next, we write the various terms in Eqn. 23 in terms of the coordinates of the chosen sticky end a that we try to bind, \mathbf{r}_a , and the distance of a from its possible binding partner b , $r_{ab} = |\mathbf{r}_b - \mathbf{r}_a|$. All other coordinates will be denoted by $\{\mathbf{r}_{\text{rest}}\}$.

The probability to be in the unbound state is then given by

$$P_a = e^{-\beta U_a} d\mathbf{r}_a r_{ab}^2 dr_{ab} d\Omega \{d\mathbf{r}_{\text{rest}}\} \frac{q_a^{\text{int}} q_b^{\text{int}}}{\Lambda_a^3 \Lambda_b^3} \quad (24)$$

where $r_{ab}^2 dr_{ab} d\Omega$ is an infinitesimal volume element around the location of b . The potential energy of the state where a is unbound is given by U_a , while q_a^{int} and q_b^{int} are the internal partition functions of a and b . Λ_a and Λ_b denote the respective de Broglie wavelengths.

When the sticky ends a and b hybridise, we place \mathbf{r}_b along the connection line \mathbf{r}_{ab} at a distance \mathcal{L} from \mathbf{r}_a . The coordinate of a remains unchanged. Since there is only one way to implement this move, the generation probability is $P_{\text{gen}}(a \rightarrow ab) = 1$.

The probability to be in the hybridised state, P_{ab} , where a and b are connected by a rod of length \mathcal{L} , is given by

$$P_{ab} = e^{-\beta U_{ab}} d\mathbf{r}_a d\Omega\{d\mathbf{r}_{\text{rest}}\} \frac{q_{ab}^{\text{int,rest}}}{\Lambda_{ab}^3}. \quad (25)$$

where U_{ab} is the potential energy of the state where a and b are hybridised. The internal partition function of the hybridised sticky ends which are restrained in their rotational freedom is denoted by $q_{ab}^{\text{int,rest}}$. Since $q_{ab}^{\text{int,rest}}$ is independent of the precise orientation of the hybridised sticky ends in space, it can be related to the internal partition of a rotationally unrestricted dsDNA segment, q_{ab}^{int} , via $q_{ab}^{\text{int,rest}} = q_{ab}^{\text{int}}/4\pi$ and thus

$$P_{ab} = e^{-\beta U_{ab}} d\mathbf{r}_a d\Omega\{d\mathbf{r}_{\text{rest}}\} \frac{q_{ab}^{\text{int}}}{4\pi\Lambda_{ab}^3}. \quad (26)$$

The factor 4π results from the fact that the connection rod between a and b is restrained from its rotational freedom. Further, Λ_{ab} is the de Broglie wavelength.

To generate the unbound state from the hybridised one, we have to generate a new coordinate for b along the connection line of the bound b and a , \mathbf{r}_{ab} . Generating the new position with a probability proportional to the distance squared, we get

$$P_{\text{gen}}(ab \rightarrow a) = \frac{r_{ab}^2 dr_{ab}}{\frac{1}{3}\mathcal{L}^3} \quad (27)$$

By imposing Eqn. 22, we get the condition for the acceptance probabilities for the Monte Carlo move:

$$\frac{P_{\text{acc}}(a \rightarrow ab)}{P_{\text{acc}}(ab \rightarrow a)} = \frac{K}{\frac{4\pi}{3}\mathcal{L}^3\rho_0} e^{-\beta\Delta U}, \quad (28)$$

with $\Delta U = U_{ab} - U_a$. In this last equation, we have used that⁷⁵

$$\frac{q_{ab}^{\text{int}}\Lambda_a^3\Lambda_b^3}{q_a^{\text{int}}q_b^{\text{int}}\Lambda_{ab}^3} = \frac{K}{\rho_0}, \quad (29)$$

with K being the equilibrium binding constant and ρ_0 the standard density of 1 mol/l.

9.2 Algorithm for the binding move in the case of several possible binding partners

Here, we outline the algorithm for the Monte Carlo binding/partner change move for several possible binding partners.

1. Choose a blob (bound or unbound) at random. Without loss of generality, we assume that this blob is on a colloid of kind A. Therefore, we denote this blob as a and its coordinates as \mathbf{r}_a .
- 2a. Find all $j = 1, \dots, N_p^u$ *unbound* blobs on unlike colloids (here: kind B), $\{b_j\}$, that are within distance \mathcal{L} , i.e. $|\mathbf{r}_{ab_j^u}| = |\mathbf{r}_{b_j^u} - \mathbf{r}_a| \leq \mathcal{L}$ (see Fig. 10a). Here, $\mathbf{r}_{b_j^u}$ is the coordinate of the unhybridised possible binding partner j .
- 2b. In case a is initially bound, add its actual binding partner to this list; the coordinate of this binding partner is denoted as $\mathbf{r}_{b_j^h}$ since it is hybridised with a . We then have a total of $N_p^u + 1$ possible binding partners.
- 3a. For all N_p^u *unbound* binding partners b_j at their unbound positions $\mathbf{r}_{b_j^u}$, do the following: along the connection line $\mathbf{r}_{ab_j^u}$, generate the new position $\mathbf{r}_{b_j^h}$ as if b_j were hybridised with a (see Fig. 10b), i.e. place it at distance \mathcal{L} from a :
$$\mathbf{r}_{b_j^h} = \mathbf{r}_a + \mathcal{L}\mathbf{r}_{ab_j^u}/|\mathbf{r}_{ab_j^u}|. \quad (30)$$
- 3b. If a was hybridised initially, further compute the following for its actual binding partner b_j : Along the connection line $\mathbf{r}_{ab_j^h}$, randomly generate a new, unbound position of b_j , $\mathbf{r}_{b_j^u}$ as

$$\mathbf{r}_{b_j^u} = \mathbf{r}_a + \sqrt[3]{x}(\mathbf{r}_{b_j^h} - \mathbf{r}_a). \quad (31)$$

with random number $x \in [0, 1)$.

After this step, both $\mathbf{r}_{ab_j^h}$ and $\mathbf{r}_{ab_j^u}$ are known for all possible binding partners.

- 4a. Calculate the weight of the state where a is not hybridised as

$$W_a = \exp(-\beta U_a), \quad (32)$$

where

$$U_a = \sum_j U_{b_j^u}. \quad (33)$$

In the last equation, $U_{b_j^u}$ is the energy that each of the j possible binding partners at their unbound positions have with the rest of the system, including the repulsion of a . Care has to be taken not to double count interactions between the various b_j .

- 4b. Calculate the weights for each of the possible hybridised states. The weight of the state where b_j is bound to a is given as

$$W_{ab_j} = \frac{K}{\frac{4\pi}{3}\mathcal{L}^3\rho_0} \exp(-\beta U_{ab_j}), \quad (34)$$

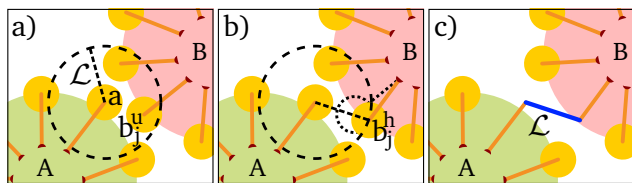


Fig. 10 (Colour online) The hybridisation move. a) From a chosen blob a on colloid A , all possible (unbound) binding partners b_j are determined which are within a sphere of radius of the hybridised sticky ends, \mathcal{L} . b) A binding partner, b_j , is chosen according to its weight (see text). This sticky end is moved from its unbound coordinates, b_j^u , to its coordinates in the hybridised state, b_j^h . c) The DNA bridge formed by the hybridised sticky ends is modelled as a rigid rod of fixed length (blue rod). Note that the figure disregards the case where a is originally bound.

where

$$U_{ab_j} = U_{ab_j^h} + \sum_{j' \neq j} U_{b_j^{u_{j'}}}, \quad (35)$$

where $U_{b_j^{u_{j'}}}$ are the energies that each of the $j' (\neq j)$ possible binding partners at their unbound positions have with the rest of the system, including the repulsion of a and b_j at their bound positions. Further, $U_{ab_j^h}$ is the interaction of b_j at its bound position with the rest of the system, including a . Again, care has to be taken not to double count interactions.

5. Calculate the sum S over all weights

$$S = W_a + \sum_j W_{ab_j}. \quad (36)$$

6. Randomly choose a state according to the probabilities of the unbound state

$$P_a = \frac{W_a}{S}, \quad (37)$$

and the various hybridised states

$$P_{ab_j} = \frac{W_{ab_j}}{S}. \quad (38)$$

If a hybridised state ab_j is chosen, a and b_j are connected by a rod and cannot move independently anymore (see Fig. 10c).

10 Appendix B: Fitting functions for effective interactions

To use effective interactions in computer simulations, it is crucial to find reliable fits for both $\Phi_{2,\text{rep}}$ and $\Phi_{2,\text{hyb}}$. The former

can be fitted by a sigmoidal curve for large distances $r > 5.5R_C$

$$\beta\Phi_{2,\text{rep}}^{\text{fit}}(r > 5.5R_C) = \frac{a_1}{1 + \exp\left(-\frac{r-b_1}{c_1}\right)}, \quad (39)$$

with fitting parameters a_1 , b_1 and c_1 . For distances $r < 5.5R_C$, the potential can be approximated by an exponential function

$$\beta\Phi_{2,\text{rep}}^{\text{fit}}(r < 5.5R_C) = a_2 \exp\left(-\frac{r}{b_2}\right) + c_2 \quad (40)$$

with fitting parameters a_2 , b_2 and c_2 . The fits have to be performed under the boundary condition that the two functions need to join smoothly at $r = 5.5R_C$.

The attractive potential $\Phi_{2,\text{hyb}}$ is fitted in two steps: for distances where DNA strands can hybridise, i.e. $r \lesssim 6R_C$, the potential is approximately linear, $\Phi_{2,\text{hyb}} \sim kx + d$. For distances $r \gtrsim 7.5R_C$, $\Phi_{2,\text{hyb}} = 0$. Then, $\Phi_{2,\text{hyb}}$ can be fit for each temperature of interest as an interpolation between these two trends⁷⁶, i.e.

$$\beta\Phi_{2,\text{hyb}}^{\text{fit}}(r; T \text{ fixed}) = kx + \tau k \ln \left[1 + \exp\left(\frac{\alpha - x}{\tau}\right) \right] + d \quad (41)$$

with $\alpha = -d/k$ and fitting parameter τ .

References

- 1 G. M. Whitesides and M. Boncheva, *Proceedings of the National Academy of Sciences*, 2002, **99**, 4769–4774.
- 2 S. C. Glotzer and M. J. Solomon, *Nature Mater.*, 2007, **6**, 557.
- 3 N. Geerts and E. Eiser, *Soft Matter*, 2010, **6**, 4647–4660.
- 4 C. A. Mirkin, R. L. Letsinger, R. C. Mucic and J. J. Storhoff, *Nature*, 1996, **382**, 607.
- 5 A. P. Alivisatos, K. P. Johnsson, X. Peng, T. E. Wilson, C. J. Loweth, M. P. Bruchez and P. G. Schultz, *Nature*, 1996, **382**, 609.
- 6 S. Y. Park, A. K. R. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz and C. A. Mirkin, *Nature*, 2008, **451**, 553.
- 7 D. Nykypanchuk, M. M. Maye, D. van der Lelie and O. Gang, *Nature*, 2008, **451**, 549.
- 8 H. Xiong, D. van der Lelie and O. Gang, *Phys. Rev. Lett.*, 2009, **102**, 015504.
- 9 R. J. Macfarlane, M. R. Jones, A. J. Senesi, K. L. Young, B. Lee, J. Wu and C. A. Mirkin, *Angew. Chem. Int. Ed.*, 2010, **49**, 4589.
- 10 M. M. Maye, M. T. Kumara, D. Nykypanchuk, W. B. Sherman and O. Gang, *Nat. Nanotechnol.*, 2010, **5**, 116.
- 11 D. Sun and O. Gang, *Journal of the American Chemical Society*, 2011, **133**, 5252–5254.
- 12 P. L. Biancaniello, A. J. Kim and J. C. Crocker, *Phys. Rev. Lett.*, 2005, **94**, 058302.

- 13 A. J. Kim, P. L. Biancaniello and J. C. Crocker, *Langmuir*, 2006, **22**, 1991–2001.
- 14 M. T. Casey, R. T. Scarlett, W. Benjamin Rogers, I. Jenkins, T. Sinno and J. C. Crocker, *Nat. Commun.*, 2012, **3**, 1209.
- 15 K. M. Ho, C. T. Chan and C. M. Soukoulis, *Phys. Rev. Lett.*, 1990, **65**, 3152–3155.
- 16 P. Cigler, A. K. R. Lytton-Jean, D. G. Anderson, M. G. Finn and S. Y. Park, *Nat. Mater.*, 2010, **9**, 918.
- 17 R. J. Macfarlane, B. Lee, M. R. Jones, N. Harris, G. C. Schatz and C. A. Mirkin, *Science*, 2011, **334**, 204–208.
- 18 Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck and D. J. Pine, *Nature*, 2012, **491**, 51.
- 19 F. Varrato, L. Di Michele, M. Belushkin, N. Dorsaz, S. H. Nathan, E. Eiser and G. Foffi, *Proceedings of the National Academy of Sciences*, 2012, **X**, X.
- 20 R. Jin, G. Wu, Z. Li, C. A. Mirkin and G. C. Schatz, *Journal of the American Chemical Society*, 2003, **125**, 1643.
- 21 R. Dreyfus, M. E. Leunissen, R. Sha, A. V. Tkachenko, N. C. Seeman, D. J. Pine and P. M. Chaikin, *Phys. Rev. E*, 2010, **81**, 041404.
- 22 A. V. Tkachenko, *Phys. Rev. Lett.*, 2002, **89**, 148303.
- 23 N. A. Licata and A. V. Tkachenko, *Phys. Rev. E*, 2006, **74**, 041408.
- 24 A. V. Tkachenko, *Phys. Rev. Lett.*, 2011, **106**, 255501.
- 25 P. Varilly, S. Angioletti-Uberti, B. M. Mognetti and D. Frenkel, *The Journal of Chemical Physics*, 2012, **137**, 094108.
- 26 F. W. Starr and F. Sciortino, *Journal of Physics: Condensed Matter*, 2006, **18**, L347.
- 27 B. Bozorgui and D. Frenkel, *Phys. Rev. Lett.*, 2008, **101**, 045701.
- 28 W. Dai, C. W. Hsu, F. Sciortino and F. W. Starr, *Langmuir*, 2010, **26**, 3601–3608.
- 29 F. J. Martinez-Veracoechea, B. Bozorgui and D. Frenkel, *Soft Matter*, 2010, **6**, 6136–6145.
- 30 F. J. Martinez-Veracoechea, B. M. Mladek, A. Tkachenko and D. Frenkel, *Phys. Rev. Lett.*, 2011, **107**, 045902.
- 31 F. Vargas Lara and F. W. Starr, *Soft Matter*, 2011, **7**, 2085–2093.
- 32 M. E. Leunissen and D. Frenkel, *J. Chem. Phys.*, 2011, **134**, 084702.
- 33 R. T. Scarlett, M. T. Ung, J. C. Crocker and T. Sinno, *Soft Matter*, 2011, **7**, 1912–1925.
- 34 C. Knorowski, S. Burleigh and A. Travesset, *Phys. Rev. Lett.*, 2011, **106**, 215501.
- 35 C. Chi, F. Vargas-Lara, A. V. Tkachenko, F. W. Starr and O. Gang, *ACS Nano*, 2012, **6**, 6793–6802.
- 36 T. I. Li, R. Sknepnek, R. J. Macfarlane, C. A. Mirkin and M. Olvera de la Cruz, *Nano Letters*, 2012, **12**, 2509–2514.
- 37 B. M. Mladek, J. Fornleitner, F. J. Martinez-Veracoechea, A. Dawid and D. Frenkel, *Phys. Rev. Lett.*, 2012, **108**, 268301.
- 38 B. M. Mognetti, M. E. Leunissen and D. Frenkel, *Soft Matter*, 2012, **8**, 2213.
- 39 N. A. Licata and A. V. Tkachenko, *Phys. Rev. E*, 2006, **74**, 040401.
- 40 S. H. Tindemans and B. M. Mulder, *Phys. Rev. E*, 2010, **82**, 021404.
- 41 S. Angioletti-Uberti, B. M. Mognetti and D. Frenkel, *Nat. Mater.*, 2012, **11**, 518.
- 42 M. E. Leunissen, R. Dreyfus, R. Sha, N. C. Seeman and P. M. Chaikin, *Journal of the American Chemical Society*, 2010, **132**, 1903–1913.
- 43 W. B. Rogers and J. C. Crocker, *Proceedings of the National Academy of Sciences*, 2011, **108**, 15687.
- 44 B. M. Mognetti, P. Varilly, S. Angioletti-Uberti, F. J. Martinez-Veracoechea, J. Dobnikar, M. E. Leunissen and D. Frenkel, *Proceedings of the National Academy of Sciences*, 2012, **109**, E378–E379.
- 45 W. B. Rogers and J. C. Crocker, *Proceedings of the National Academy of Sciences*, 2012, **109**, E380.
- 46 O.-S. Lee and G. C. Schatz, *The Journal of Physical Chemistry C*, 2009, **113**, 2316–2321.
- 47 V. A. Ngo, R. K. Kalia, A. Nakano and P. Vashishta, *J. Phys. Chem. C*, 2012, **116**, 19579.
- 48 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, *The Journal of Chemical Physics*, 2011, **134**, 085101.
- 49 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, *Phys. Rev. Lett.*, 2010, **104**, 178101.
- 50 Y. Zhang, H. Zhou and Z.-C. Ou-Yang, *Biophys. J.*, 2001, **81**, 1133.
- 51 S. J. Hurst, A. K. R. Lytton-Jean and C. A. Mirkin, *Analytical Chemistry*, 2006, **78**, 8313–8318.
- 52 D. Nykypanchuk, private communication.
- 53 L. J. Henderson, *Am. J. Physiol.*, 1908, **21**, 173.
- 54 S. V. Kuznetsov, Y. Shen, A. S. Benight and A. Ansari, *Biophys. J.*, 2001, **81**, 2864.
- 55 B. Tinland, A. Pluen, J. Sturm and G. Weill, *Macromolecules*, 1997, **30**, 5763.
- 56 S. B. Smith, Y. Cui and C. Bustamante, *Science*, 1996, **271**, 795.
- 57 B. M. Mognetti *et al.*, to be published.
- 58 J. B. Mills, E. Vacano and P. J. Hagerman, *Journal of Molecular Biology*, 1999, **285**, 245 – 257.
- 59 D. Frenkel and B. Smit, *Understanding Molecular Simulation*, Academic Press, London, 2nd edn, 2002.
- 60 B. M. Mladek and D. Frenkel, *Soft Matter*, 2011, **7**, 1450.
- 61 P. G. Bolhuis, A. A. Louis, J.-P. Hansen and E. J. Meijer,

-
- The Journal of Chemical Physics*, 2001, **114**, 4296.
- 62 N. R. Markham and M. Zuker, *Nucleic Acids Res.*, 2005, **33**, W577.
- 63 R. A. Van Santen, *J. Phys. Chem.*, 1984, **88**, 5768.
- 64 J. H. Holland, *Adaptation in Natural and Artificial System*, The University of Michigan Press: Ann Arbor, 1975.
- 65 D. Gottwald, G. Kahl and C. N. Likos, *J. Chem. Phys.*, 2005, **122**, 204503.
- 66 J. Fornleitner, F. Lo Verso, G. Kahl and C. N. Likos, *Soft Matter*, 2008, **4**, 480.
- 67 J. Fornleitner, F. Lo Verso, G. Kahl and C. N. Likos, *Langmuir*, 2009, **25**, 7836.
- 68 G. J. Pauschenwein, *J. Phys. A: Math. Theor.*, 2009, **42**, 355204.
- 69 B. M. Mladek, P. Charbonneau and D. Frenkel, *Phys. Rev. Lett.*, 2007, **99**, 235702.
- 70 B. M. Mladek, P. Charbonneau, C. N. Likos, D. Frenkel and G. Kahl, *Journal of Physics: Condensed Matter*, 2008, **20**, 494245.
- 71 M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, ninth Dover printing, tenth GPO printing edn, 1964.
- 72 E. J. Meijer, D. Frenkel, R. A. LeSar and A. J. C. Ladd, *J. Chem. Phys.*, 1990, **92**, 7570.
- 73 M. von Smoluchowski, *Z. Physik. Chem.*, 1917, **92**, 129–168.
- 74 *Diffusion-Limited Reactions*, ed. C. H. Bamford, C. Tipper and R. G. Compton, Elsevier, 1985, vol. 25, pp. 3 – 46.
- 75 B. Bozorgui, *PhD thesis*, University of Amsterdam, 2008.
- 76 C. C. Kerr, S. J. van Albada, C. J. Rennie and P. A. Robinson, *Clinical Neurophysiology*, 2010, **121**, 962 – 976.